Least-to-Most Prompting Enables Complex Reasoning in Large Language Models

Year: 2022 | *Citations: 1427* | *Authors: Denny Zhou, Nathanael Scharli, Le Hou*

Abstract

Chain-of-thought prompting has demonstrated remarkable performance on various natural language reasoning tasks. However, it tends to perform poorly on tasks which requires solving problems harder than the exemplars shown in the prompts. To overcome this challenge of easy-to-hard generalization, we propose a novel prompting strategy, least-to-most prompting. The key idea in this strategy is to break down a complex problem into a series of simpler subproblems and then solve them in sequence. Solving each subproblem is facilitated by the answers to previously solved subproblems. Our experimental results on tasks related to symbolic manipulation, compositional generalization, and math reasoning reveal that least-to-most prompting is capable of generalizing to more difficult problems than those seen in the prompts. A notable finding is that when the GPT-3 code-davinci-002 model is used with least-to-most prompting, it can solve the compositional generalization benchmark SCAN in any split (including length split) with an accuracy of at least 99% using just 14 exemplars, compared to only 16% accuracy with chain-of-thought prompting. This is particularly noteworthy because neural-symbolic models in the literature that specialize in solving SCAN are trained on the entire training set containing over 15,000 examples. We have included prompts for all the tasks in the Appendix.